(2)

STATE OF THE ART OF NATURAL LANGUAGE PROCESSING

Name of Contractor: The Futures Group
76 Eastern Boulevard
Glastonbury CT 06033-1264

Sponsored by

Defense Advanced Research Projects Agency (DoD)

Defense Small Business Innovation Research Program

ARPA Order No. 5916

Issued by U.S. Army Missile Command Under

Contract #DAAH01-87-C-0750

Phone Number: (203) 633-3501

Effective Date of Contract: 08 July 1987

Contract Expiration Date: 15 Nov 1987

Reporting Period: 15 Nov 1987

Principal Investigator: Thomas Anderson

CONTENTS

Abstract	v
INTRODUCTION	1
BACKGROUND	3
METHODOLOGY	7
NATURAL LANGUAGE PROCESSING INTERVIEW PROTOCOL	9
INTERVIEW SUMMARY	11
GRAPHICAL RESULTS	25
ALTERNATIVE PERSPECTIVE	35
OVERALL CONCLUSIONS	37
Appendices	
A. Letter of Introduction	39
B. The Futures Group "State of the Art" Measurement Methodology	43
Glossary	47

ABSTRACT

A study was carried out to determine the "state of the art" of the natural language processing requirements of a battle management system. The study was based on a methodology developed by this contractor ~~The Futures Group. The results of the~~ study results indicate the field is in an early stage of development and further progress will be required to achieve the tools for a natural language interface to a battle management system. Key words

Keywords: Natural Language Processing, natural language, artificial intelligence, Information Processing, battle management, "state of the art"

INTRODUCTION

This study was undertaken to determine whether the SOA methodology developed by The Futures Group could be useful in defining the "state of the art" of the Natural Language Processing domain of Artificial Intelligence. The study was carried out under an SBIR contract for DARPA (DAAH01-87-C-0750) between July 8, 1987, and November 15, 1987. The study is based on structured interviews with six experts in the field of Natural Language Processing. The experts were chosen from a list supplied by DARPA. The interviews were held at the facilities of the expert interviewees.

The study was an outgrowth of prior work in technology assessment by The Futures Group. In the prior studies, technologies such as microprocessors, gas turbine engines, batteries and other high-technology components were analyzed using the SOA methodology. The methodology was also used to access the SOA of computer languages. The results of these studies provided numerical evaluations of the SOA of the subject areas and were intuitively satisfying to individuals who were experts in those fields. Experts in the field were aware of many of the nuances that the SOA methodology could not deal with; however, it was generally agreed that the thrust of the field was captured. The results were in a form readily understood by someone who was not an expert in the field.

This study is the first time the methodology was applied to a field whose products were primarily laboratory studies. This presented a problem in that the "state of the art" is generally thought of from a product point of view. What this study attempts to demonstrate is that the SOA of the component technology is necessary to construct a product which contains "Natural Language Processing." It is really a measure of the "state of the art" of the tools necessary to build an Artificial Intelligence system with Natural Language Processing capabilities.

BACKGROUND

Natural Language Processing is that portion of the field of Artificial Intelligence that is devoted to attempting to understand how people use language, with the goal of capturing that capability with a machine. There are a number of reasons why Natural Language Processing capabilities are highly desirable. These include machine translation of text, natural language interfaces to computer programs and machines that can understand speech.

The primary difference between a natural language such as English and an artificial language (programming language) is the avoidance of ambiguity in the artificial language by having a highly structured syntax. Digital computers were first programmed by inputting binary code into the machine to represent instructions and data. The process was tedious and error prone. To simplify the process, artificial programming languages were created. The highly structured syntax made programming easier but required programmers to learn a new language. These new programming languages required the user to spend large amounts of time learning the language and severely constrained the way information was input and output to the computer. It was evident to the earliest users of computers that a machine which understood natural languages was highly desirable. It also was evident from the work of Chomsky* and others that the science of linguistics or understanding of human language was inadequate to serve as a basis for Natural Language Processing as applied to computers.

---

*Noam Chomsky, Aspects of the Theory of Syntax (Cambridge, Mass.: MIT Press, 1965).

One of the earliest attempts at Natural Language Processing was the attempt to perform machine translation of textual materials. This work was first carried out in the Soviet Union and shortly afterward in the United States. Charniak and McDermott in their book, Introduction to Artificial Intelligence,* devoted two pages to the early attempts in the United States to perform machine translation. They entitled the section, "The Sad Story of Machine Translation." They showed that the science and engineering basis for Natural Language Processing which was available at the time was inadequate to the task. At the present time, we still have only a primitive understanding of how people use language and the mechanism by which they understand. The great technological revolution experienced by the electronics industry had as its basis a firm understanding of solid-state physics. The field of Natural Language Processing may require a similar scientific foundation for it to gain the widespread use forecast for it.

The work carried out in this study of the State of the Art of Natural Language Processing by The Futures Group was an attempt to quantify the scientific basis for Natural Language Processing for a particular application. The application chosen was battle management, which is an important application of Natural Language Processing. The methodology is based on work performed for the National Science Foundation. It has been extensively applied to hard technologies such as computers, microcomputers, batteries and other devices. In addition, we performed a study for the Department of Defense on computer languages using the same methodology. This was the first time we have attempted to apply this methodology to a body of knowledge rather than a product.

It was not our intention at the outset to study the scientific basis for Natural Language Processing; however, it quickly became evident that few products were

---

*E. Charniak and D. McDermott, Introduction to Artificial Intelligence, (Reading, Mass.: Addison-Wesley, 1986).

available and their recent introductions would not form a basis for understanding the history of the field. The historic input is of paramount importance for analyzing the state of the art of a subject because it is a measure of performance that changes with time. For these reasons, we chose to use laboratory programs that had limited objectives as the basis for the study. This complicates the analysis because each program had limited objectives and did not incorporate all the capabilities that a Natural Language Processing product might have incorporated. This tends to understate the capabilities of the field at any particular time. It is not, however, meaningless because most Natural Language Processing programs built on previous work tend to incorporate a significant number of the features of their predecessors. In addition, Artificial Intelligence programs that use Natural Language Processing only incorporate that amount of Natural Language Processing necessary to perform the task.

METHODOLOGY

Six experts from five institutions were interviewed for this study. Two of the interviewees were from academic institutions and four were from the research departments of commercial firms. All had at least 10 years' experience in AI Natural Language Processing, and the average experience level was closer to 20 years. All the interviewees were educated to the Ph.D. level and most had extensively published in AI literature. The interviewees were evenly divided between East and West Coast institutions. All the interviewees were actively engaged in Natural Language Processing research.

We attempted to interview eight individuals. Two were unavailable. We believe the results were not altered due to interviewing six rather than eight individuals.

The interviewees were contacted by letter (see Appendix A) with follow-up via telephone. The respondents were interviewed at their respective facilities. The interview protocol (see page 9) was designed for a one-and-one-half-hour length interview. All the respondents were generous with their time and the interviews were actually 2 to 3 hours in duration. Anonymity was guaranteed to each of the respondents so that an unencumbered response could be obtained. In return for their cooperation, we stated that we would make the results of the study available and answer any future questions that might arise from this effort.

NATURAL LANGUAGE PROCESSING

INTERVIEW PROTOCOL

Question 1. Has the field of Natural Language Processing improved over the last 10-20 years?

Question 2. Do you have a model of what you believe represents the operation of Natural Language Processing?

Question 3. What are the measures of performance that would indicate progress had taken place?

Question 4. Will you identify for us specific products or programs that represent major steps in the history of Natural Language Processing?

Question 5. On a scale of 1-5, can you rate the performance criteria identified in Question 3 for each of the programs identified in Question 4?

<u>Performance Scale</u>

1. Able to define problem
2. Limited understanding
3. Limited useful applications
4. Widespread application
5. Complete understanding of subject

Question 6. How would you rate the performance requirements for a battle management application of Natural Language Processing?

Question 7. Is speech a driving force in Natural Language Processing?

Question 8. Can you identify U.S. centers of excellence and individuals who you believe are at the cutting edge of Natural Language Processing research?

Question 9. Are there centers of excellence outside the United States that are driving Natural Language Processing?

Question 10. Where do you believe the field is going in the next 5-10 years?

Question 11. Can the field of Natural Language Processing be described by a level model as shown below?

Lexicon — The kind of information found in dictionaries: the definitions of the word and its word class.

Syntax — The structure form of sentences.

Semantics	The meaning of the sentence with respect to the text or dialog in which it is contained.
Pragmatics	The domain knowledge required to make sense of the words or sentences.
Learning	The ability to incorporate new knowledge into the program based on discourse or interaction with the program.

The "state-of-the-art" analysis methodology developed by The Futures Group requires an application to delineate the performance parameters. The application we chose was a battle management program that we modeled as a large interactive data base, an expert system capable of interacting with the data base, and a natural language processing interface for the user. We described the user as an aircraft carrier-based force commander operating in an area such as the Persian Gulf. The system was presumed to work with a Yeoman typing information into a console, with the information read from a monitor. What we sought with this application was to go beyond the limited-domain natural language interfaces to highly constrained data bases.

INTERVIEW SUMMARY

Question 1

Has the field of Natural Language Processing improved over the last 10-20 years?

A. Work in labs has progressed over the last 10 years but applications are still being built using 1970s technology.

B. Improvements in semantic and syntactic processing are aiding speech recognition.

C. Major shifts in the field and considerable improvement in tools have occurred. Semantic understanding is greater.

D. Yes, progress was made in formalizing English grammar, and in syntax and semantics in the 1970s. We have come a long way in languages representing meaning, pragmatics, and discourse. Learning is a fuzzy area not yet deeply researched.

E. Definite progress, but you cannot measure that progress.

F. Progress has been made in:

Theoretics	Research structures of language, syntax, pragmatics Computational aspects--algorithms of syntax and semantics Language processing in interactive discourse
Technology	Customization of research for commercial applications Good systems are available

CONCLUSION

Yes, there is improvement. Each participant described improvement in terms peculiar to personal experiences and applications. There was minimal correlation between various responses to this question.

Question 2

Do you have a model of what you believe represents the operation of natural language processing?

A. & B. A series of boxes

C. ----------

D. See Figure 1

E. ----------

F. See Figure 2

CONCLUSION

We were able to find only one coherent model of the process. Nevertheless, even this model was described by the particular respondent as incomplete and highly subject to change.

We proposed a highly simplistic model consisting of a common bus and processing elements, which were connected (Figure 1). The response was that yes, this model might represent some aspects of Natural Language Processing, but was by no means a workable model.

Our conclusion is that a lack of one or many cohesive models prevents a truly quantitative assessment of the functional components of Natural Language Processing. In at least one case, there was complete disagreement as to what were the functional components.

Question 3

What are the measures of performance that would indicate progress has taken place?

A. & B.
- Syntax area is best understood of all "boxes"
- Abstract grammars
- Context-free languages
- Acceptance of natural languages
- Linguistic-syntactic formalisms that can be adapted to natural languages
- Adapted to natural languages
- Augmented context-free grammar
- Unification grammars
- Processors and compilers . · unification languages
- Move from sentence efforts to extended discourse
- How to deal with what users really mean
- Branches--connective graph provers, Prolog, technical theorem provers

C.
- Multi-language concepts
- Computational frameworks
- Transformational parsers
- Chart parsing
- General rewriting system
- Moving away from procedurality
- Doing it by characterization of structures
- Partitioning space

D.
- Lexicon
- Syntax
- Semantics
- Discourse (Dialog)
- Pragmatics
- Learning

E.
- Conceptual information
- Inferencing
- Memory indexing
- Memory
- Reminding, and many other topics

F.
- Lexicon
- Syntax
- Semantics
- Discourse
- Pragmatics

CONCLUSION

A lack of a coherent model prevented the majority of respondents from rating performance on the basis of functional components.

<u>The absence of uniformly acceptable models may be a reason why measurable performances criteria are generally felt to be unattainable.</u>

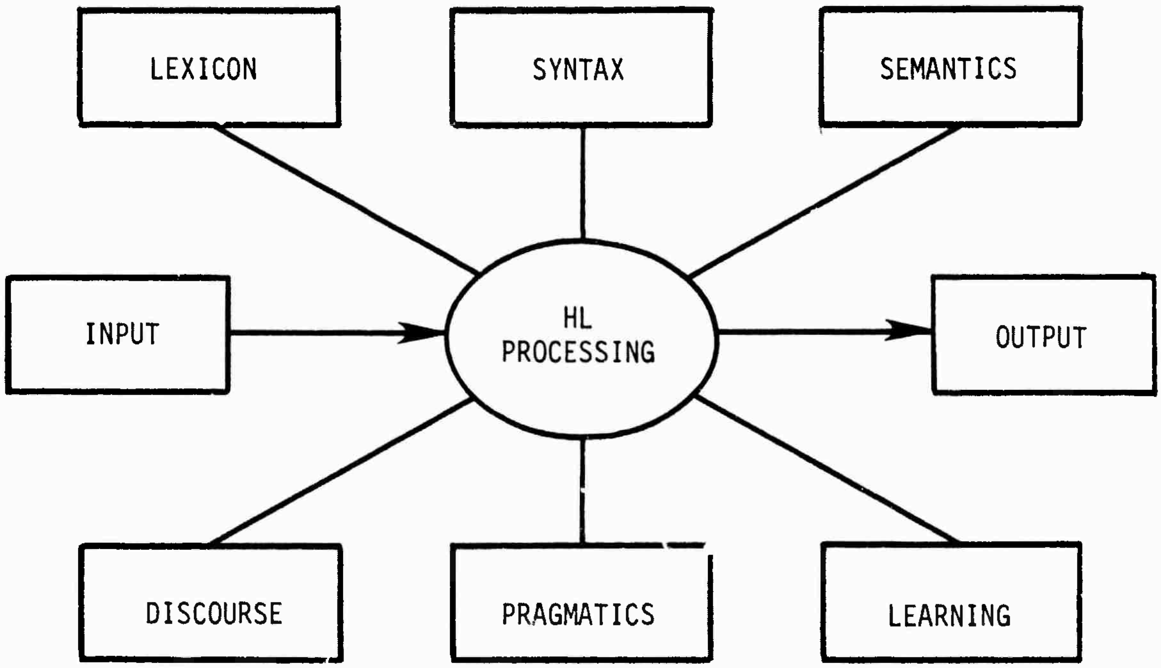
LEXICON
SYNTAX
SEMANTICS
INPUT
HL
PROCESSING
OUTPUT
DISCOURSE
PRAGMATICS
LEARNING

Figure 1

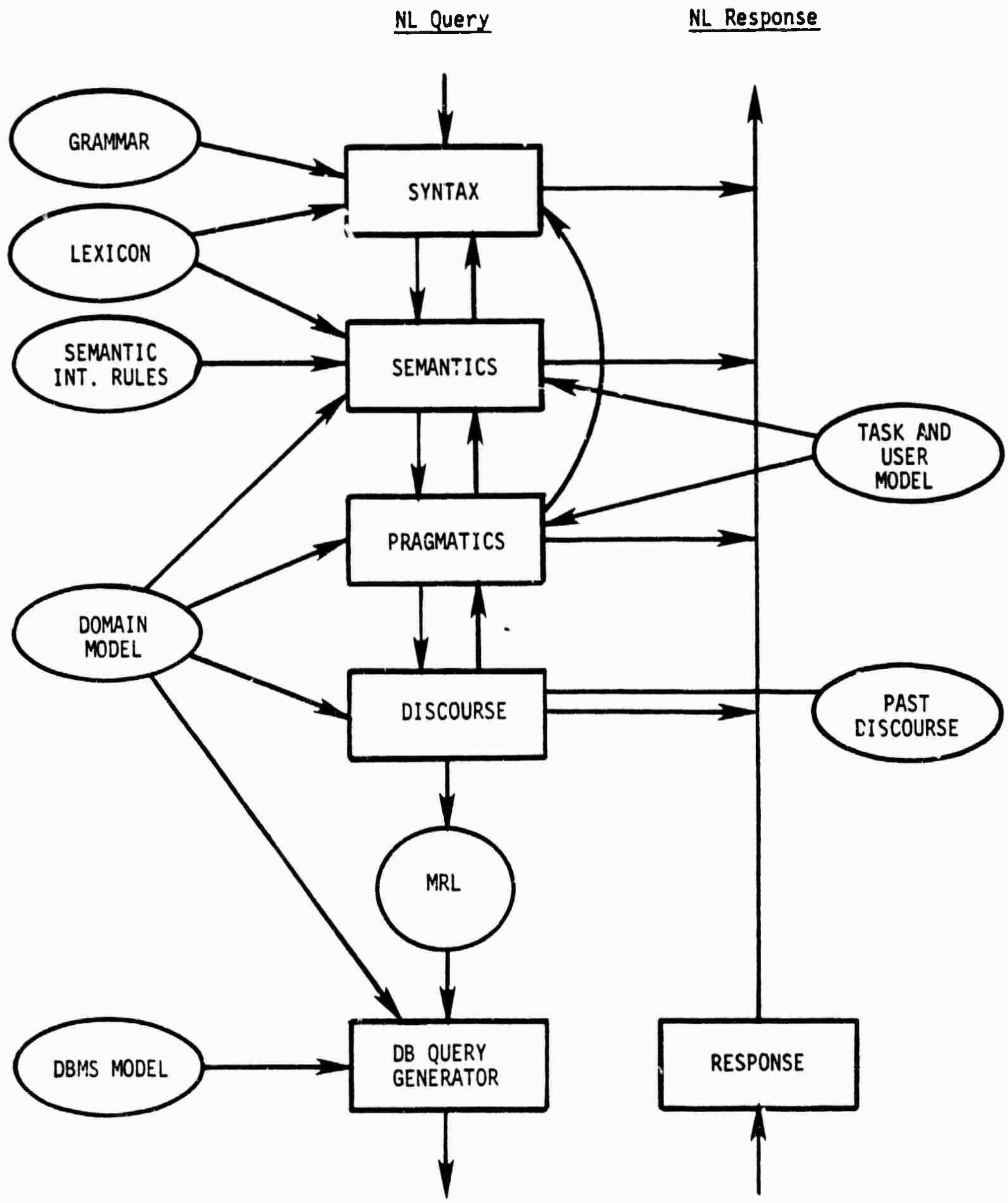
NL Query
NL Response
GRAMMAR
LEXICON
SEMANTIC INT. RULES
SYNTAX
SEMANTICS
PRAGMATICS
DISCOURSE
TASK AND USER MODEL
DOMAIN MODEL
PAST DISCOURSE
MRL
DBMS MODEL
DB QUERY GENERATOR
RESPONSE

Figure 2

Question 4

Will you identify for us products or programs that you are familiar with that represent major steps in the history of Natural Language Processing?

A. & B. Parlance, Data Talker, Clout, McDonald Douglas System "with outrageous claims," Micro-mini-English.

C. LFG, work of Chomsky, Hewlett-Packard, Generalized Phase Structure Grammar.

D. Lunar, DARPA speech understanding, Schank's Conceptual Dependency Theory.

E. ----

F. Text, CoOp, Romper, Spirit, IRAS, Mumble, RTM, Grumble.

CONCLUSION

All respondents have a different perception of the milestones achieved in the history of the field. Some perceptions may be colored because the firm is involved in specific commercial development programs.

A telling indication may be the widely diverse backgrounds that are represented by various researchers (computer science, psychology, philosophy, linguistics, anthropology, "artificial intelligence," and/or some combination of the above). This diversity is institutionalized by the equally diverse departmental structures found in companies and major universities engaged in AI research.

This is yet another factor mitigating against a uniform perception of model(s) of the field of Natural Language Processing.

Question 5

On a scale of 1-5, can you rate the performance criteria for each of the programs identified in Question 4?

Results are shown in Figures 3 through 8.

CONCLUSION

We were unable to obtain measurable performance criteria from two-thirds of the respondents. Half the respondents were "unable" to delineate measurable performance criteria. One respondent did not believe that measurable performance criteria were either important or relevant.

Those interviewees who responded rated the performance criteria based on a performance scale proposed by us but acceptable to the interviewees. The performance evaluations we arrived at, and their scaling, are shown in Tables 1-7 and Figures 3-8.

Question 6

How would you rate the performance requirements for a battle management system?

This information is an integral part of the SOA methodology. It sets out the specific requirements of a task rather than the overall performance for the field. The "state-of-the-art" definition we use in our methodology requires a specific task.

The average criteria given for the present time are:

Performance Criteria	Scale (1-5)
Lexicon	4
Syntax	4
Semantics	3
Discourse	2
Pragmatics	2
Learning*	2

*One of the respondents was not sure that learning was well enough defined to be a performance criterion.

CONCLUSION

Again, because of the aforementioned constraints, the ability to assess performance requirements for a theoretical battle management system was limited.

Question 7

Is speech a driving force in Natural Language Processing?

A. & B. Speech is a tool that is part of Natural Language Processing systems. The existence of speech understanding will have a major influence on commercial acceptance of Natural Language Processing.

C. Not important at present time, will be in future.

D. Speech is nice but not important.

E. Speech is a frill.

F. Speech is a necessary part of Natural Language Processing systems.

CONCLUSION

We found the respondents to have widely varying opinions as to the role of speech recognition and generation. The spectrum of opinion ranged from speech topics as being essential to being insignificant in the overall progress of Natural Language Processing.

Question 8

Can you identify U.S. centers of excellence and individuals who you believe are at the cutting edge of Natural Language Processing research?

A. & B. See Proceedings of Computational Linguistics of July 4, 1987, Stanford University, Stanford; MIT; Roger Schank, Yale; Terry Winograd, Stanford; SRI.

C. Carnegie-Mellon; Xerox; Berkeley.

D. Hewlett-Packard; Ray Perrault, SRI; James Allen, University of Rochester; Don Walker, Belicore; Barbara Gross, Harvard.

E. Yale; Jamie Carbonell, Carnegie-Mellon; Jerry Young, University of Illinois; Chris Hammond, University of Chicago; Wendy Leonard.

F. University of Pennsylvania; BBN, Inc.

CONCLUSION

As in the answers given in Question 7, there was no uniform consensus. Again, this reflects the dearth of agreed-upon goals and the means to achieve them.

Question 9

Are there centers of excellence outside the United States?

A. & B. Eurotran and Japanese effort.

C. Stuttgart, Germany (translations of LFG applications); IKOT Japan, plus other major Japanese companies.

D. Eurotran is too early to tell.

E. European effort is a joke. There is some significant work being done in Canada.

CONCLUSION

The "Eurotran" project was referenced by most of the respondents. The range of responses varied from "serious effort" to "it is a joke." One respondent felt that work in Canada was significant. Another identified some important work in progress in Japan. No one indicated that major work in Natural Language Processing would be achieved outside the United States.

Question 10

Where do you believe the field is going in the next 10 years?

A. & B. Next generation of commercial systems will have extended discourse capability. Achievement of speech recognition in next two years will have major influence on development of new commercial systems.

C. Dramatic improvements in capabilities will be achieved in 2-to-5-year time span.

D. Draining of resources as limited applications are achieved. Need statements of problem issues. Commercial systems will stick pretty much to semantics and syntax.

E. We have been uncovering the layers and we think we may be seeing the final layer. Perhaps major new discoveries in the next 2-5 years.

F. Sense of optimism that the next 2-5 years will see major progress in systems. Integration of Natural Language Processing and graphics. Multi-modal systems--speech, graphics, Natural Language Processing.

CONCLUSION

There was a surprising uniformity of belief that the field will undergo major advances over the next two to five years. This is amazing in light of the lack of common perceptions in the approaches to the field. Each respondent had different specific reasons why there would be advances in the overall field of Natural Language Processing.

Question 11

Can the field of natural language processing be described by a level model (lexicon, syntax, semantics, discourse, pragmatics, learning)?

A. & B. Not asked.

C. Not asked.

D. Yes, with learning added to original list.

E. No, totally inappropriate to the science of Natural Language Processing, although I recognize others in the field would accept that breakdown.

F. Yes, that is a generally accepted breakdown. However, learning is not well defined by research at the present time.

CONCLUSION

This question was introduced halfway through the study. The reason for its introduction, with reservations, was lack of uniformity in the identification of performance criteria.

One of the interviewees rejected the criteria as totally inappropriate. The remaining interviewees accepted our breakdown as adequate (with reservations) to categorize the "building blocks" of Natural Language Processing.

GRAPHICAL RESULTS

1. The graphical results of lexicon performance indicate a relatively mature subject in 1979 and small incremental growth over the next 10 years.

2. The results for syntax are similar to the results for lexicon. This building block is in a moderately mature state and progress is expected to be incremental.

3. The progress in semantics is probably understated by the results of Figure 5. The results indicate incremental changes in the future. However, progress in both discourse and pragmatics will necessarily determine the pace of advancement in semantics.

4. Figure 6 shows discourse to be in an early stage of development with little improvement over the past 6-8 years.

5. Figure 7 shows pragmatics to be in an early stage of development with performance being difficult to ascertain.

6. The low overall performance for Natural Language Processing with a battle management application is surprising. This could indicate that performance criteria in question do not form the basis for a battle management system of the scope required.

LEXICON

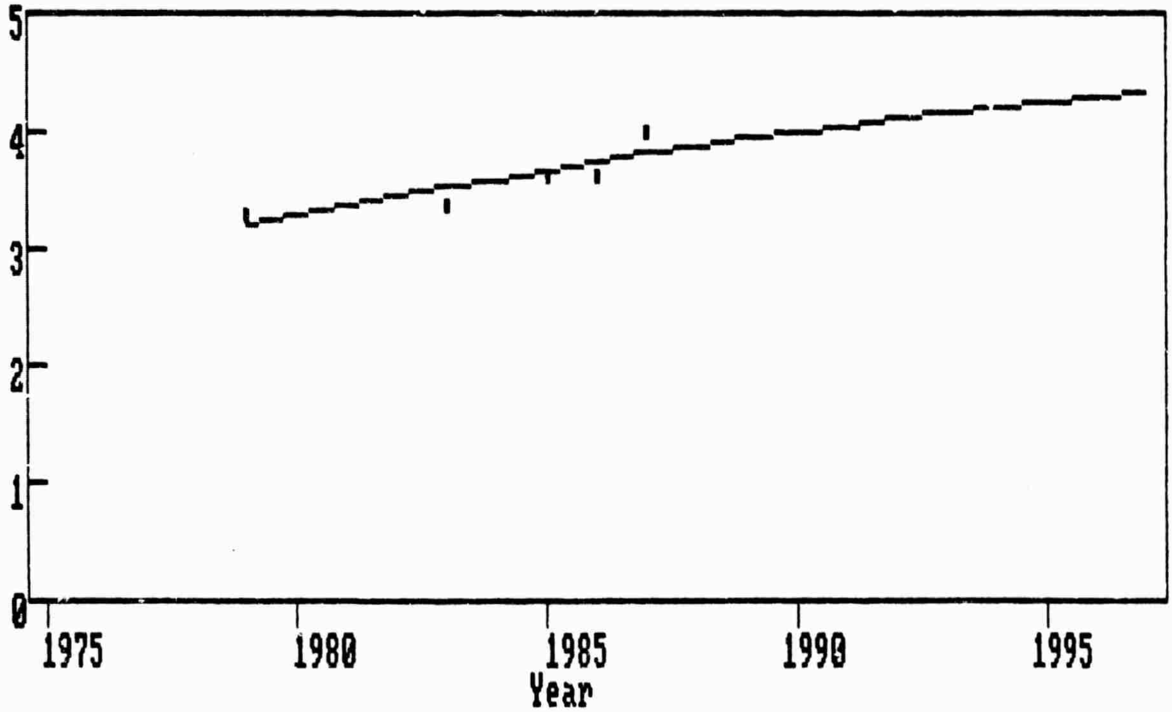

Figure 3

Table 1

LEXICON

Technology	Year	Var Given	Var Fit	Error
LAB PRGM	1979	3.31000	3.20169	0.10831
LAB PRGM	1983	3.39000	3.52630	-0.13630
LAB PRGM	1985	3.61000	3.67515	-0.06515
LAB PRGM	1986	3.63000	3.74587	-0.11587
LAB PRGM	1987	4.00000	3.81402	0.18598

Variable forecast based on historical S-shape curve : LEXICON

Year	VAR Forecast
1988	3.87959
1989	3.94253
1990	4.00284
1991	4.06054
1992	4.11563
1993	4.16816
1994	4.21815
1995	4.26568
1996	4.31078
1997	4.35354

SYNTAX

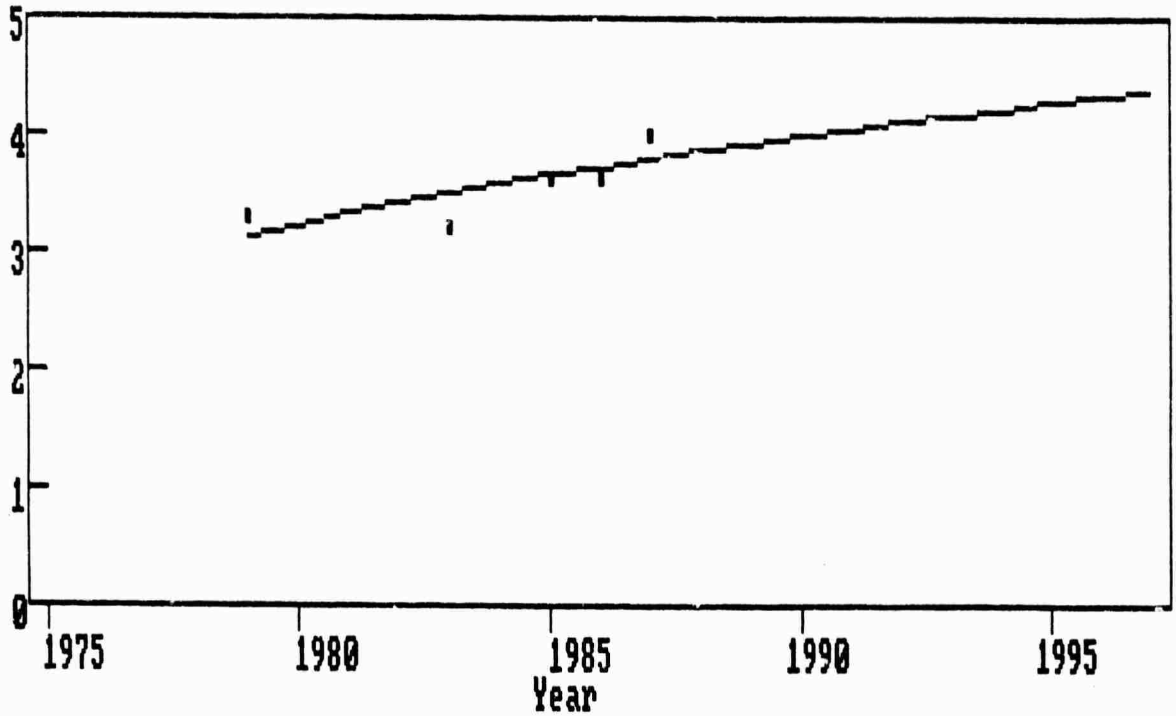

Figure 4

Table 2

SYNTAX

Technology	Year	Var Given	Var Fit	Error
LAB PRGM	1979	3.31000	3.13788	0.17212
LAB PRGM	1983	3.20000	3.48642	-0.28642
LAB PRGM	1985	3.61000	3.64611	-0.03611
LAB PRGM	1986	3.63000	3.72187	-0.09187
LAB PRGM	1987	4.00000	3.79478	0.20522

Variable forecast based on historical S-shape curve : SYNTAX

Year	VAR Forecast
1988	3.86481
1989	3.93191
1990	3.99607
1991	4.05731
1992	4.11563
1993	4.17109
1994	4.22372
1995	4.27359
1996	4.32078
1997	4.36535

SEMANTICS

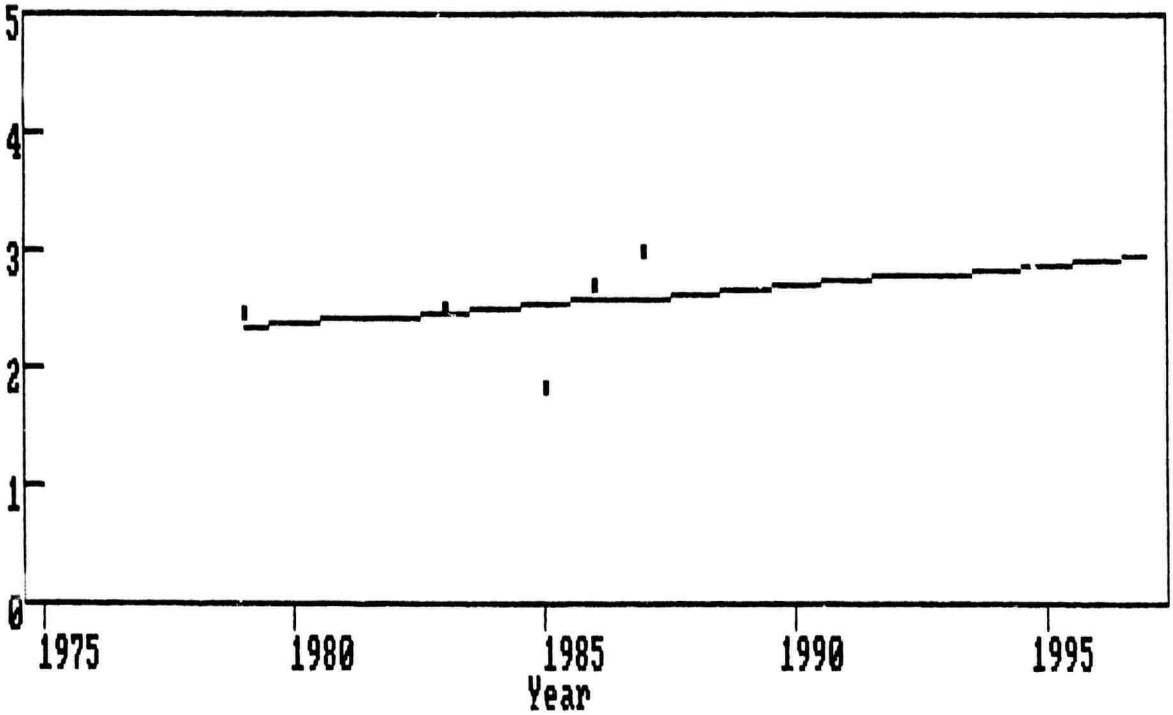

Figure 5

Table 3

SEMANTICS

Technology	Year	Var Given	Var Fit	Error
LAB PRGM	1979	2.47000	2.32724	0.14276
LAB PRGM	1983	2.51000	2.46536	0.04464
LAB PRGM	1985	1.82000	2.53455	-0.71455
LAB PRGM	1986	2.71000	2.56913	0.14087
LAB PRGM	1987	3.00000	2.60369	0.39631

Variable forecast based on historical S-shape curve : SEMANTICS

Year	VAR Forecast
1988	2.63821
1989	2.67267
1990	2.70707
1991	2.74139
1992	2.77562
1993	2.80974
1994	2.84375
1995	2.87762
1996	2.91136
1997	2.94494

DISCOURSE

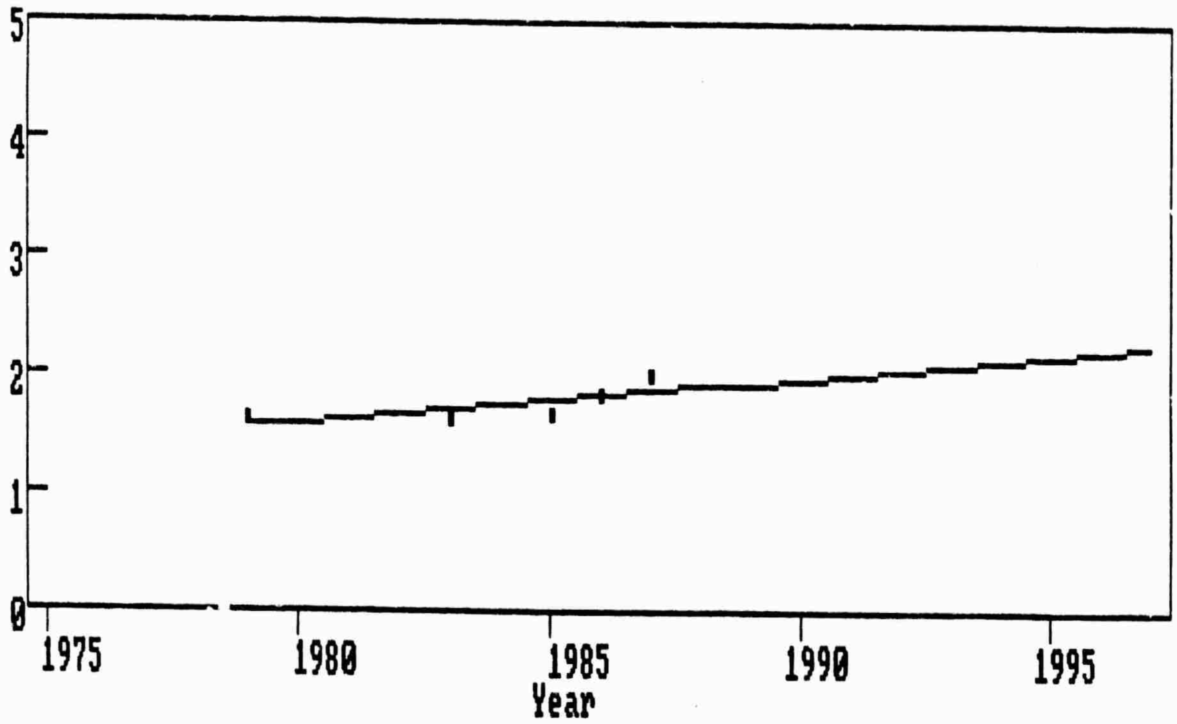

Figure 6

Table 4

DISCOURSE

Technology	Year	Var Given	Var Fit	Error
LAB PRGM	1979	1.63000	1.56451	0.06549
LAB PRGM	1983	1.64000	1.70846	-0.06846
LAB PRGM	1985	1.65000	1.78277	-0.13277
LAB PRGM	1986	1.82000	1.82047	-0.00047
LAB PRGM	1987	2.00000	1.85849	0.14151

Variable forecast based on historical S-shape curve : DISCOURS

Year	VAR Forecast
1988	1.89684
1989	1.93549
1990	1.97443
1991	2.01364
1992	2.05310
1993	2.09278
1994	2.13268
1995	2.17278
1996	2.21304
1997	2.25346

PRAGMATICS

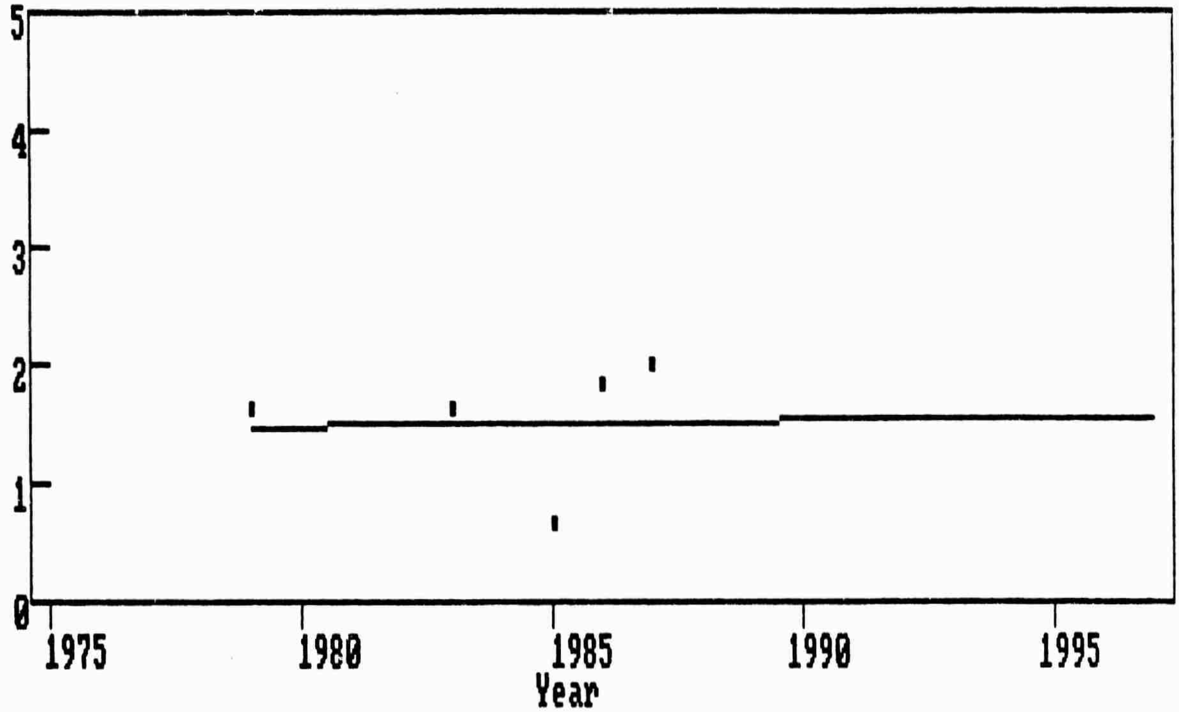

Figure 7

Table 5

PRAGMATICS

Technology	Year	Var Given	Var Fit	Error
LAB PRGM	1979	1.63000	1.47405	0.15595
LAB PRGM	1983	1.64000	1.49152	0.14848
LAB PRGM	1985	0.68000	1.50030	-0.82030
LAB PRGM	1986	1.82000	1.50470	0.31530
LAB PRGM	1987	2.00000	1.50911	0.49089

Variable forecast based on historical S-shape curve : PRAGMTCS

Year	VAR Forecast
1988	1.51352
1989	1.51795
1990	1.52238
1991	1.52682
1992	1.53126
1993	1.53571
1994	1.54017
1995	1.54464
1996	1.54911
1997	1.55359

LEARNING

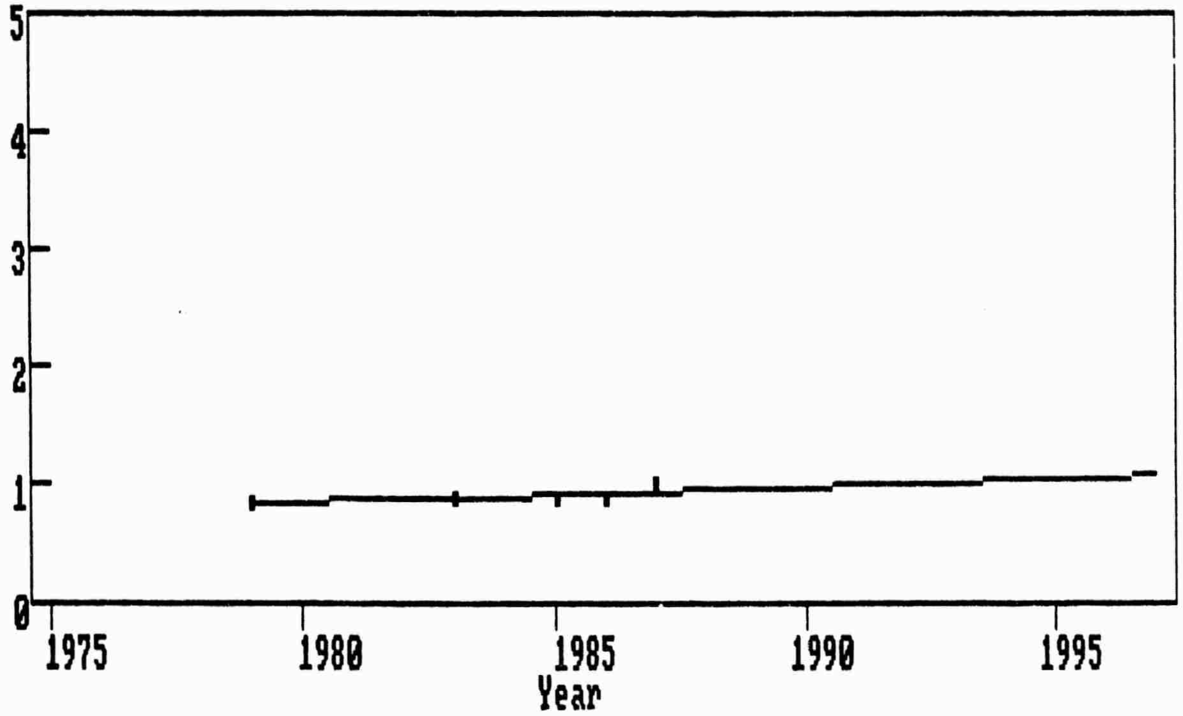

Figure 8

Table 6

LEARNING

Technology	Year	Var Given	Var Fit	Error
LAB PRGM	1979	0.85000	0.83079	0.01921
LAB PRGM	1983	0.87000	0.88008	-0.01008
LAB PRGM	1985	0.86000	0.90558	-0.04558
LAB PRGM	1986	0.89000	0.91854	-0.02854
LAB PRGM	1987	1.00000	0.93165	0.06835

Variable forecast based on historical S-shape curve : LEARNNG

Year	VAR Forecast
1988	0.94490
1989	0.95829
1990	0.97183
1991	0.98552
1992	0.99934
1993	1.01332
1994	1.02743
1995	1.04170
1996	1.05611
1997	1.07066

Table 7

NATURAL LANGUAGE PROCESSING

SOA Solution

Variables:		Weights:	Maximum:
Variable 1	LEXICON	0.160	5.000
Variable 2	SYNTAX	0.160	5.000
Variable 3	SEMANTICS	0.180	5.000
Variable 4	DISCOURS	0.160	5.000
Variable 5	PRAGMTCS	0.180	5.000
Variable 6	LEARNNG	0.160	5.000

Given Data:

Technology	Year	Var1	Var2	Var3	Var4	Var5	Var6
LAB PRGM	1979	3.31	3.31	2.47	1.63	1.63	0.85
LAB PRGM	1983	3.39	3.20	2.51	1.64	1.64	0.87
LAB PRGM	1985	3.61	3.61	1.82	1.65	0.68	0.86
LAB PRGM	1986	3.63	3.63	2.71	1.82	1.82	0.89
LAB PRGM	1987	4.00	4.00	3.00	2.00	2.00	1.00

Computed Data:

Technology	Year	SOA Computed	SOA Fit	SOA Error
LAB PRGM	1979	0.43880	0.41839	0.02041
LAB PRGM	1983	0.44060	0.45057	-0.00997
LAB PRGM	1985	0.40136	0.46684	-0.06548
LAB PRGM	1986	0.48212	0.47501	0.00711
LAB PRGM	1987	0.53200	0.48318	0.04882

Forecast based on S-shape extrapolation of variables:

Year	Average SOA Forecast	Upper Frontier Forecast	Lower Frontier Forecast
1988	0.48822	0.50563	0.47182
1989	0.49545	0.51312	0.47880
1990	0.50251	0.52043	0.48563
1991	0.50940	0.52757	0.49229
1992	0.51613	0.53453	0.49879
1993	0.52269	0.54133	0.50513
1994	0.52908	0.54796	0.51131
1995	0.53532	0.55441	0.51734
1996	0.54140	0.56071	0.52322
1997	0.54732	0.56685	0.52894

ALTERNATIVE PERSPECTIVE

An alternative view of the entire field of Natural Language Processing exists, wherein the attempt to adapt a quantitative assessment is a difficult undertaking at best. Whereas the experts previously discussed are concerned with a "building-block" approach toward integrating Natural Language Processing, an alternative approach entirely discounts this method. Rather, this approach emphasizes a much broader concern with enabling a machine understanding of "stories." Instead of the part-task breakdown of lexicon, syntax, semantics, discourse, pragmatics, etc., this view finds limited use in that approach.

A clear distinction is made between advances in scientific or laboratory research, and engineering or commercial applications. As with most technological advances, there is a "freeze" of scientific advances when translated into engineering application. Commercial applications of natural language research presently are utilizing advances that are perhaps ten years old or older. Of course, the limited engineering applications of Natural Language Processing are what provide grist for the unwarranted popular notion that "thinking" computers are imminent.

The alternative approach to Natural Language Processing scientific research addresses the nature of understanding and seeks to progress through a spectrum of such for application in computers. This spectrum begins with an area called "making sense," through "cognitive understanding," and ultimately toward "complete empathy." At issue is research into enabling a computer to learn from mistakes (Schank*). This field pursues enabling computers to "explain things to

---

*R. C. Schank, Explanation Patterns (Hillsdale, New Jersey: Lawrence Erkbaum Associates, 1986).

themselves, to question experience, and to be creative in explanation not previously expressed." Throughout the research a key topic is memory (and memory organization) and its use.

This alternative approach is exceedingly more complex in its scope and aims than is the building-block method. It seeks to probe at the very root of human understanding to be able to translate this into methods usuable for computers. The leaders in this field steadfastly refuse to quantify gains in Natural Language Processing in the terms used by the building-block linguistic approach. Rather, scientific research gains can only be measurable in terms of task orientation. Progress has been made--but only in terms of progress over where the field of Natural Language Processing was "X" amount of years earlier. Engineering applications are almost disdained insomuch as they detract resources from the ultimate goals of scientific research. Instead of using an arithmetic means of quantifying progress, the method seeks to establish a means of computer understanding of myriads of stories for progression toward "cognitive understanding" and beyond. Intense efforts exist in relating a spectrum of needs to different levels of explanation. Short-term breakthroughs are neither sought nor desired. Work in the area of explanation patterns and all its implications seems to be at the heart of this effort in Natural Language Processing. As advances are made in these areas toward the postulated core problems, the feeling is that the peripheral issues will be filled in concurrently or after-the-fact. No timetables are established.

Despite the assertion that mainly scientific, vice-engineering advances are important, there have been significant commercial applications of Natural Language Processing products from this group of experts. Hence, although advances have been achieved and will most assuredly continue to do so under this approach, they are not easily measurable in terms used elsewhere in this examination of Natural Language Processing state of the art.

OVERALL CONCLUSIONS

Determinations of the overall state of the art of Natural Language Processing is a difficult undertaking due to the inherent nature of the field. Nevertheless, given a group of parameters that were agreed upon, with reservation, by some of the researchers, it was possible to apply SOA methodology to "building block" components of Natural Language Processing. Bear in mind that some experts do not agree that any measure of progress has meaning, or is even possible. The problems do not seem to lie in the methodology, but in the overall lack of cohesive structure in some researchers' Natural Language Processing efforts. Given these reservations, the SOA of some components of Natural Language Processing are:

- Lexicon. The subject matter is moderately well understood. Changes and future improvements are expected to be incremental. Major problems to be resolved include word ambiguities, which are also addressed in syntax, semantics, and discourse.

- Syntax. Syntaxes which can be parsed are moderately well developed. Again, future improvements are expected to be incremental.

- Semantics. Considerable progress has been made in the past 10 years, particularly in understanding context-free sentences. Problems exist in understanding conjunctions, in understanding quantifiers, and in understanding negations. Semantics is being driven by research in progress in discourse and pragmatics.

- Discourse. This area is in an early stage of development. Problems remain with the issues of ellipsis and anaphora. In battle management programming, ellipsis may become a major issue if speech recognition is a requirement. Anaphora is more of an issue in document translation and text translation.

- Pragmatics. Also in an early development stage, this area has been described as being the "dumping ground" for problems not dealt with elsewhere. The pragmatic requirements vary inversely with the breadth of the domain of the subject. The domain of battle management will specify the amount of pragmatics expertise required for a Natural Language Processing interface.

- Learning. This area is poorly defined at present. With one notable exception, researchers are not heavily oriented in this direction yet. We believe this to be a major requirement of any future Natural Language Processing system. This area will be a crucial aspect of a successful battle management system.

Appendix A

LETTER OF INTRODUCTION

Sepbember 21, 1987

Dear

I am doing research in the field of Natural Languages, specifically attempting to measure the "State of the Art" of Natural Language tools. This work is being sponsored by DARPA, and the Program Manager, Lt. Col. Robert Simpson, Ph.D. (Program Manager, Machine Intelligence), recommended we interview you as one of the major contributors to the field. Our "state-of-the-art" procedure is based on analysis of historic information as well as analysis of presently available software. The heart of the analysis is interviews with experts such as yourself. We ask the experts to tell us what they believe performance criteria for the subject matter should be. Using those performance criteria, we then ask them to rate presently available tools and historic software programs they are knowledgeable about.

The performance criteria and the rating of historic programs are done in structured interviews that take approximately one and one-half hours. The interviewee is treated as a confidential source of information and is not identified by name or organization in our report. We may reference the person by numbers of publications of books, etc.

The procedure, which we call "state of the art," was developed by The Futures Group under National Science Foundation sponsorship. It has been extensively used to measure the "state of the art" of hard technology (microprocessors, super-computers, and memory chips), and we have some experience applying it to software (computer languages, and operating systems). I have enclosed a copy of a paper describing the technique. We believe attempting to measure the "state of the art" of a field as complex as natural languages may be audacious. We have, however, found that using a broad-brush technique such as "state of the art" in a complex field sometimes produces a degree of clarity that is absent when all the nuances are accounted for.

I would like to schedule an interview with you and any of your colleagues that you feel are knowledgeable about present and historic programs in "natural languages." I plan to schedule the interviews for the week of October 12-16, 1987. I will attempt to contact you by phone early in October 1987.

Very truly yours,

Thomas M. Anderson
Senior Scientist

TMA:gjl
Enclosure

Appendix B

THE FUTURES GROUP "STATE-OF-THE-ART" MEASUREMENT METHODOLOGY

THE FUTURES GROUP "STATE-OF-THE-ART" MEASUREMENT METHODOLOGY

The Method

Although "state of the art" is a familiar term, it lacks precision. Generally, when a technology is described as state of the art, it is taken as an example of an advanced development in the field, but there is no way of indicating the degree of advancement. The Futures Group has developed a convention for measuring technology state of the art utilizing an index comprised of selected performance parameters (or variables) that describe a particular technology. The approach has proved versatile in its ability to capture technological performance at various levels of system aggregation and in relating increases in state of the art to developments in component technologies and advances in design. By plotting the state-of-the-art indicator for each new product/innovation over time, the path of technological development can be quantitatively described.

In this convention, the state of the art (SOA) of a particular product or process is defined as a linear combination of a number of factors or parameters descriptive of that product or process. While non-linear equation forms can be used, the basic function form of the state-of-the-art is as follows:

$$SOA = K_1(P_1/P'_1) + K_2(P_2/P'_2) \cdots K_n(P_n/P'_n)$$

where n is the number of parameters that are taken to define the technology, P_n is the value of the n^{th} parameter, P'_n is a reference value of the n^{th} parameter (used to nondimensionalize the equation), and K_n is the weight--that is, the relative importance of the n^{th} parameter.

This equation defines progress in the technology as improvements in the parameters selected to describe the technology. The relative contribution of each of the parameters to the overall state of the art is determined by the selected weights. Selection of the parameters which describe the technology and the weights with which these parameters will be applied are key issues, and both judgmental and statistical methods are available for parameter and weight selection.

To use the method, an analyst must first define the intent of the technology. This emphasis on use is important since we believe that the state of the art of any technology depends upon how well it fulfills its design purposes.

There are, in general, two approaches to determining specifically which parameters to include and their associated weights; these are expert judgment and statistical methods. In this study, expert judgments are relied on to select parameters and their weights. In this approach, experts are asked to provide their judgments about the list of factors important to the performance definition of a particular technology and to assign weights to each of the factors.

GLOSSARY

Lexicon	The level of language that deals with definitions of words and word classes.
Syntax	The arrangement of order of words in a sentence. The structure of a sentence.
Semantics	The study of the meanings of sentences.
Discourse	A level of language understanding that derives meaning from multi-sentence analysis. It deals with the problem of resolving sentence ambiguity by finding meaning in context with other sentences.
Pragmatics	The level of language understanding that incorporates domain knowledge to derive meaning in a discourse. Meaning is inferred from common knowledge relating to scripts, goals, and common activities rather than the specific words or sentences in the discourse.
Learning	The level of Natural Language Processing that attempts to incorporate new knowledge into a system. It assumes the Natural Language Processing system understands input and modifies its behavior accordingly.
Ellipsis	The omission of a word or words from a sentence.
Anaphora	The problem of dealing with abbreviations in a Natural Language Processing system.
Natural Language	Any of the commonly used languages--English, French, Spanish, etc.
Artificial Language	Computer programming languages such as Fortran, Basic, Pascal, Lisp and Prolog.
Parser	A computer program capable of syntactically breaking down a sentence.
Domain	The total body of knowledge required to understand a subject.